University of Groningen

Emodiversity

Brown, Nicholas J.L.; Coyne, James C.

Published in:
Journal of Experimental Psychology. General

DOI:
10.1037/xge0000330

Emodiversity: Robust Predictor of Outcomes or Statistical Artifact?

Nicholas J. L. Brown

James C. Coyne

University of Groningen

Acknowledgements

The authors wish to thank Toon Kuppens, Neuroskeptic, Thomas Paternoster-Howe, Jaap Scherphuis, and Sam Schwarzkopf for their helpful comments on an earlier version of this article, and Jordi Quoidbach for kindly and promptly supplying the data set for Study 1. Any errors in this article are the responsibility of the authors alone.

Abstract

This article examines the concept of emodiversity, put forward by Quoidbach et al. (2014) as a novel source of information about "the health of the human emotional ecosystem" (p. 2057). Quoidbach et al. drew an analogy between emodiversity as a desirable property of a person's emotional make-up and biological diversity as a desirable property of an ecosystem. They claimed that emodiversity was an independent predictor of better mental and physical health outcomes in two large-scale studies. Here, we show that Quoidbach et al.'s construct of emodiversity suffers from several theoretical and practical deficiencies, which make these authors' use of Shannon's (1948) entropy formula to measure emodiversity highly questionable. Our reanalysis of Quoidbach et al.'s two studies shows that the apparently substantial effects that these authors reported are likely due to a failure to conduct appropriate hierarchical regression in one case, and to suppression effects in the other. It appears that Quoidbach et al.'s claims about emodiversity may reduce to little more than a set of computational and statistical artifacts.

Quoidbach et al. (2014) presented a novel construct, *emodiversity*, to represent the degree to which an individual experiences a wide range of emotions, analogous to the idea of biodiversity in the natural environment. Emodiversity can be conceptualized either on separate axes of positive and negative emodiversity, or as a single overall “global emodiversity.” In two large studies, Quoidbach et al. claimed to have found associations between emodiversity and health outcomes, independent of the associations between the corresponding positive or negative emotions and those outcomes: “Emodiversity is a practically important and previously unidentified metric for assessing the health of the human emotional ecosystem” (Quoidbach et al., 2014, p. 2057).

Quoidbach et al. (2014) operationalized emodiversity in terms of a concept from the field of information theory known as *Shannon entropy* (Shannon, 1948). They claimed that this concept, originally devised by Shannon to represent the information content of a message in a communication system, also “quantifies the number of species and the evenness of species in a biological ecosystem” (p. 2058). In support of this claim, they cited Magurran (2004), who actually cautioned that “most commentators who discuss the relative merits of the various methods of measuring diversity go out of their way to underline the disadvantages of the Shannon index” (Magurran, 2004, p. 106). Leaving aside that debate, however, the validity of Quoidbach et al.’s concept clearly depends on the degree to which it faithfully implements Shannon entropy; its ability to generate reliable, meaningful, and measureable variance across participants; and the behavior of emodiversity when applied to real data. The present article examines whether emodiversity lives up to these requirements.

This article is structured as follows. First, we examine the theoretical underpinnings of emodiversity, especially the applicability of Shannon entropy to the specific context of a multi-item measure of emotions using Likert-type responses. Second, we reanalyze Quoidbach et al.’s (2014) empirical findings to identify where the purported evidence for their remarkable claims about the health benefits of emodiversity might have come from. Finally, we briefly discuss the broader lessons to be learned from this case.

Theoretical issues

Limitations of the analogy with biodiversity

Biodiversity is defined in terms of the *richness* and *evenness* of the variety of species within an ecosystem. Richness is the number of distinct species to be found in a given sample, regardless of how many examples (provided that the number is greater than zero) of that species are detected, while evenness is the degree to which the populations of each species (or, in some definitions, the corresponding biomass) are similar. In Quoidbach et al.'s (2014) definition of emodiversity, the equivalent of biological richness is the number of different emotions experienced, while the equivalent of evenness is the extent to which the frequency with which a person reports experiencing each emotion is similar across all of the measured emotions. However, both of these dimensions of emodiversity are subject to limitations that are not present in Shannon's (1948) original model or in the Shannon-Wiener index of biodiversity. These limitations severely impair the correspondence between emodiversity and Shannon entropy.

Richness. In a biodiversity setting, the richness of a community is usually unbounded. Certainly, it would be unusual for richness to be subjected to an a priori upper limit imposed by the designer of a field study. With emodiversity, however, the degree to which the richness of participants' emotional experience can be captured is limited by the number of items making up the measure being used. For example, in their Study 1, Quoidbach et al. (2014) used the mDES scale, asking participants how often they experienced each of nine positive and nine negative emotions in a recent period. In contrast, the PANAS-X (Watson & Clark, 1994) measures 30 different positive emotions (and 30 negative ones); for example, this scale allows participants to provide distinct reports of the extent to which they feel *calm*, *relaxed*, and *at ease*. Leaving aside for a moment questions of the meaning of taxonomies of emotions (e.g., Kristjánsson, 2003; Solomon, 2002) and the possible discrepancies—whether caused by demand characteristics, social desirability concerns, faulty recollection, or some other issue—between people's actual emotional experiences and how they describe them, especially if they are asked about experiences in the past rather than the present (Robinson & Clore, 2002), this would seem to imply that emodiversity should be measured using a scale that allows the greatest possible number of emotions to be reported. By analogy, sending a field biologist out to report the number of (only) rabbits, mice, rats, voles, and beavers in a given area, while ignoring foxes or wolves because there was no corresponding space on the form, might lead to suboptimal decisions about conservation policy. However, Quoidbach et al. did not address this question, although it appears to be of crucial theoretical importance; in their two empirical studies, they used measures with only nine and 10 positive and negative emotions, respectively.

Evenness. When the Shannon-Wiener index is used in field biology, the count associated with each species is precisely the number of examples of that species observed in the community. However, a self-report by participants in a psychological study of the extent to which they experienced various emotions will not typically be expressed in terms of a count of discrete occasions. A question such as "How many times did you experience contentment in the last week?" is almost meaningless; for one thing, the existence of 20 distinct occasions of contentment implies 20 transitions from a state of non-contentment, so that somebody who remained contented the whole time would have a very low score. Like most researchers asking about past emotional states, Quoidbach et al. (2014) used measures that asked participants to report the relative frequency with which they experienced each emotion on a Likert-type scale, with the following possible responses: 0: *Never*; 1: *Rarely*; 2:

Some of the time; 3: *Often*; 4: *Most of the time*[1]. But this decision imposes severe limitations on the potentially measurable (un)evenness of emotions (to go with the limitations noted above on the measurement of richness), because the range of possible values is so small. It is equivalent to a field biologist who uses the Shannon-Wiener index to measure biodiversity reporting the numbers of each species that were observed as *None*, *A few*, *Some*, *Quite a lot*, and *Very many*, with these groupings subsequently being coded 0–4 and the resulting numbers injected into Shannon's formula. Such an operation would result in the loss of a great deal of information compared to the recording of the exact numbers of each species that were observed.

Could this problem be alleviated by asking participants to report the frequency with which they experienced each emotion on a wider numerical scale, perhaps in the range 0–100? This would seem to allow the relative frequency of emotional experiences to be described with greater (albeit still bounded) precision, thus allowing for more differentiation of the "evenness" component of emodiversity. However, it is unclear whether many individuals are sufficiently aware of their emotional experiences that they could meaningfully assign values of, say, 34 to one and 37 to another. In a classic paper, Miller (1956) described the general problem of people's limited ability to distinguish between more than about seven levels of a unidimensional construct. Experience from several fields with these kinds of numerical-rating scales (e.g., Berbaum, Dorfman, Franken, & Caldwell, 2002; DeSoto, 2016; Mickes, Wixted, & Wais, 2007) suggests that the most responses will likely be multiples of

---

[1] In fact, Quoidbach et al. (2014) did not report the exact responses corresponding to the values 1, 2, and 3. We have assumed that the version of the mDES that these authors used in Study 1 was similar to that described by Fredrickson (2013), from which the responses shown here are taken, and that the responses for the unspecified 10-item measure used in Study 2 were labeled similarly.

10, with a tendency to cluster further around 0, 50, and 100. The problem for emodiversity is that if each respondent only uses, say, five different numbers from the range of 0–100 to describe their frequencies of emotional experience, then the result is mathematically equivalent to the case where they are constrained to choose from a range of 0–4, and this holds even if every participant chooses a different set of five numbers. Perhaps a better way to measure the range and variety of a person's emotional experiences, suggested by a reviewer of the present article, might be to use some form of diary, with sampling of emotions at either regular or random intervals (e.g., the Experience Sampling Method; Csikszentmihalyi & Larson, 1987); provided that a sufficient number of "species" of emotion could be defined, this could potentially result in measures of evenness that are more mathematically meaningful.

Numerical effects of bounded richness and evenness

Quoidbach et al.'s (2014) failure to correctly implement the concept of Shannon entropy, described in the preceding section, has immediate and severe consequences for the numerical behavior of the measurement of emodiversity. We enumerated all possible combinations of zero (*Never*) and non-zero responses for a variety of possible numbers of scale items and response formats (0–4 and 0–100). The results, shown in detail in our Supplemental Information, demonstrate that unless a participant responds that they *Never* experience about two-thirds of the emotions being measured—a contingency that would probably raise questions about the validity of the instrument being used—the ratio between the lowest and highest possible emodiversity values is very small (about 1.1:1 for a scale having a 0–4 response format and 1.5:1 for one with a 0–100 response format). Furthermore, even this range of values is only possible under rather implausible circumstances, because the highest and lowest scores for emodiversity are obtained when participants exhibit highly unusual response patterns (all identical responses for the maximum score; alternating extreme responses, or large numbers of *Never* responses, for the minimum score). Normally, however, psychologists do not say that individuals who display such response patterns "have high [or low] emotional diversity." Rather, they say that such people "are not paying attention," and typically exclude these participants' data altogether. Hence, the practical range of emodiversity values from any given measure of emotions is even less than the mathematical limits would suggest. To summarize, we believe that emodiversity—as defined by Quoidbach et al. using the Shannon entropy formula and implemented using short emotion measures with limited ranges of responses—is unlikely to provide any meaningful amount of variance, independent of the underlying emotion measure, to be explained empirically.

Abundance is not measured

Quoidbach et al. (2014) defined people's emodiversity as "the variety and relative abundance of the emotions they experience" (p. 2057). However, it is not clear that the formula these authors used to calculate emodiversity fully corresponds to this definition. The formula generates the highest possible value of emodiversity when all of the responses are equal and non-zero: "If *all* the emotions of the list were evenly experienced, then emodiversity would be maximal" (Quoidbach et al., 2014, p. 2059; emphasis in original). However, this maximum value of emodiversity is attained whatever this non-zero value might be (1, 2, 3, or 4). This means that Alice, who experiences all nine positive emotions (alertness, amusement, etc.) only *Rarely*, has exactly the same emodiversity as Bob, who experiences all of these emotions *Most of the time*, and higher emodiversity than Carol, who experiences four of the nine emotions *Some of the time* and five of them *Often*. Furthermore, any increase in Alice's experience of one of these emotions will result in an immediate *reduction* in her emodiversity. Given Quoidbach et al.'s claims about the health benefits of higher emodiversity, this would appear to be a good reason for Alice not to attempt to increase her currently minimal frequency of, say, amusement to *Some of the time* by watching comedy shows until she is sure that she can simultaneously increase the frequency with which she experiences the other positive emotions to the same extent.

Thus, once a minimum level of frequency of experiencing emotions has been established, there can be no "benefit" (of increased emodiversity) in increasing the frequency with which one experiences any particular emotion, unless this specific emotion is in some way "lagging" behind the others (for example, if one currently experiences eight emotions *Often* and one only *Some of the time*). It is therefore not clear where the "relative abundance" component of Quoidbach et al.'s (2014) definition of emodiversity is to be found, given that a person experiencing all positive emotions only *Rarely* already has the highest possible emodiversity score. This result also suggests that any difference in emodiversity between two participants who report *Never* experiencing the same number of emotions is very likely to be nothing more than noise.

Empirical Issues

In the first part of this article, we have shown that Quoidbach et al. (2014)'s emodiversity is merely a pastiche of Shannon entropy, with a number of conceptual lacunae that make the application of Shannon's (1948) formula invalid. (For reasons of space, we have omitted some other important problems, such as the apparent requirement that participants be forced to provide responses to all items on the emotion measures being used, or the questionable validity of applying the arithmetic of Shannon's formula to numbers that are merely categorical labels on participants' reports of the frequency with which emotions were experienced; these and other issues are covered in detail in our Supplemental Information.) In fact, we believe that any observed variance in emodiversity is likely to be little more than statistical noise. In view of this, it is necessary to explain how Quoidbach et al. (2014) concluded that emodiversity independently predicted depression (Study 1) or an assortment of physical health outcomes (Study 2) to the substantial degree that they reported in their article. In this part, therefore, we examine how these results might have come about, if this were not due to the claimed predictive power of emodiversity.

Study 1

Dr. Quoidbach (personal communication, November 15, 2015) kindly provided us with the data set for Study 1, which examined the relation between emodiversity and depressive symptoms, while controlling for emotions, in a sample of 35,844 participants who were recruited via a television show. Using SPSS, we were able to reproduce perfectly all of the results reported by Quoidbach et al. (2014) in their article. However, we conducted the majority of our reanalyses in R; our code is available at https://osf.io/vu4uq/.

Quoidbach et al. (2014) claimed that the results of their regressions showed that emodiversity substantially and significantly predicted their main outcome variable (depression), over and above the effect of positive or negative emotions themselves. For example, when positive emotion, positive emodiversity, and their interaction were entered into a regression, the reported standardized regression coefficients (β) of these three terms were, respectively, −.40, −.36, and −.14[2]. However, simply reporting these coefficients (and their corresponding partial *r*s) does not demonstrate a substantial effect of emodiversity. First, as our Table 1 shows, when positive emodiversity and the interaction term are entered

[2] When we reproduced this regression in SPSS, we obtained the same results reported by Quoidbach et al. (2014). However, when we performed the same operations in R, the β coefficient for the interaction was −.09, although the other two coefficients were identical. We wonder whether this might not already be an indication that multicollinearity may be causing problems for the OLS regression algorithms used by one or both of these software packages.

into a hierarchical regression after first entering positive emotion, there is almost no increase in the variance explained at each step; R^2 increases from .381 (positive emotion only) to .394 (addition of positive emodiversity) to .397 (addition of the interaction term). That is, the addition of emodiversity and the interaction term make very little difference to the explanatory power of the model, with just 1.3% of extra variance explained by emodiversity and another 0.3% explained by the interaction. (For negative emodiversity, the amount of additional variance explained is essentially zero, as explained in the section entitled "The effect of adding quadratic terms to the regression models" in our Supplemental Information.) Second, when the interaction term is added, the variance inflation factors (VIFs) of all three terms become quite large, further suggesting that multicollinearity is playing a major role (as could be expected from the correlation of .75 between positive emotion and positive emodiversity). Third, when the interaction term is entered, the standardized regression coefficient (β) for emodiversity more than doubles in magnitude, from −.175 to −.364, suggesting that substantial confounding effects are emerging at this point.

Our Table 1 also shows the similar results that are obtained when hierarchical techniques are used to explore the other regression analyses in Quoidbach et al.'s (2014) Study 1. For negative emotion and emodiversity, R^2 remains static (within the limitations of rounding) at .460 when emodiversity is added to the basic regression of depressive symptoms on negative emotion, and increases only to .466 when the interaction term is added; furthermore, similar patterns of high VIFs and suppression values are observed as for positive emotion and emodiversity. The addition of global emodiversity to a model predicting depression from positive and negative emotion results in a negligible increase in R^2 (from .592 to .595), and further suppression effects can be observed. Taken together, these results suggest that emodiversity has little independent contribution to make to the prediction of depressive symptoms, over and above the well-established role of positive and negative emotions in that regard.

Study 2

In their Study 2, using a nationally representative sample of 10,000 Belgian adults, Quoidbach et al. (2014) claimed that both positive and negative emodiversity, when entered into a multiple regression together with the corresponding emotion scores and the interaction between these two variables, predicted health outcomes such as doctor's visits, days spent in hospital, and medical expenditure better than factors whose associations with health are well-established, such as diet, exercise, and smoking. If these claims were to be verified, the implications for public health policy would be substantial. In principle, a simple questionnaire asking people how often they experienced particular emotions in the past week would potentially provide better information about the population's health status for a whole year than complex measures of actual behavior.

Unfortunately, the data for Quoidbach et al.'s (2014) Study 2 are not currently publicly available. As a result, we have been unable to determine whether the same issues concerning the lack of additional predicted variance when emodiversity is added to the regression model, seen in Study 1, are also present in Study 2. However, even without access to the data, it can be readily shown that many of the results in Study 2 are the result of statistical suppression effects. For example, when Quoidbach et al. entered positive emodiversity, mean positive emotion, and their interaction into a regression predicting doctor's visits, they reported obtaining a standardized regression coefficient (β) for positive emodiversity of −.29. But their Table 2 shows that the zero-order correlation between positive emodiversity and doctor's visits was just −.05. The presence, in multiple regression results, of a β coefficient of greater magnitude (or different sign) to the zero-order correlation between the same variables indicates that suppression has occurred. Suppression is often seen in regression models when two predictors are correlated, with at least one of them having a correlation with the outcome variable that is either zero (a situation referred to as "classical" suppression) or only relatively modest in magnitude compared to the correlation of the other predictor ("negative suppression"). However, suppression will also (always) occur when two predictors are positively correlated with each other—even to a very small extent—while having correlations of opposite sign with the outcome ("reciprocal suppression").

Inspection of Quoidbach et al.'s (2014) Table 3 suggests that suppression effects are behind the great majority of the results from Study 2. Of the 18 zero-order correlations between positive, negative, and global emodiversity and Quoidbach et al.'s list of six health outcomes, only two (positive emodiversity correlated with doctor's costs, and negative emodiversity correlated with hospital costs, both at $r = -.07$) are large enough to be statistically significant at the .05 level, even with the sample size of 1,273; this suggests that there is uncertainty about the magnitude, and even the sign, of some of these correlations. However, when these data were used in a regression model, Quoidbach et al. reported standardized regression coefficients (β) of greater magnitude than the zero-order correlations in all 18 cases, with 16 of these being statistically significant at the .05 level. Each of these coefficients appears to be the product of suppression effects. For positive emodiversity, these effects mostly take the form of negative suppression, resulting from the high correlation between positive emotion and emodiversity, combined with the fact that the zero-order correlations between positive emodiversity and the outcome variable are smaller than those between positive emotion and the outcome variable[3]. In contrast, for negative emodiversity

[3] Specifically, for two predictors X_1 and X_2 and an outcome Y, suppression will occur if the pattern of correlations between these three variables is such that $rYX_1/rYX_2 < rX_1X_2$, where

the suppression effects mostly take the form of reciprocal suppression, caused by the positive correlations between negative emotion and emodiversity and between negative emotion and the outcome, combined with the negative correlation between negative emodiversity and the outcome (Conger, 1974).

In the absence of a solid theoretical explanation, results based on suppression are typically uninterpretable. Paulhus, Robins, Trzesniewski, and Tracy (2004) presented what they claimed were two reproducible examples of suppression situations in psychology, but there do not seem to be many major effects in the psychological literature that are consistently and reliably explained in terms of theoretically-justified suppression. In particular, the existence of classical, negative, and reciprocal suppression effects in the same study seems difficult to explain theoretically, since these effects result from different patterns of relationships between predictors and outcome (Conger, 1974). It seems to us that by far the most parsimonious explanation here is that most of the statistically significant β coefficients in Quoidbach et al.'s (2014) Study 2 are the result of a combination of noisy data and the high correlation between the predictors (i.e., positive or negative emotion and the corresponding form of emodiversity).

Conclusion

Quoidbach et al. (2014) drew some far-reaching conclusions about the impact of emodiversity on mental and physical health from their two correlational studies. From their Study 1, they concluded that emodiversity has incremental predictive validity, over and above that associated with emotions of the same valence, for depression. From their Study 2, they concluded that emodiversity was at least as good a predictor of physical health as regular

---

rAB is the zero-order correlation between any two variables A and B, and the identifiers X_1 and X_2 are assigned so that $rYX_1 < rYX_2$.

exercise, a healthy diet, and refraining from smoking. However, as we have shown, these results are almost certainly nothing more than a statistical mirage. Once the regression analyses in Study 1 were conducted in a hierarchical manner it became clear that the incremental effect of emodiversity in terms of added variance explained was negligible, while in Study 2 the apparently substantial regression coefficients were shown to be the product of suppression effects. In both cases, the fact that most of these regression coefficients and their associated partial correlation coefficients were statistically significant at the traditional .05 level was neither surprising (given the large sample sizes used by Quoidbach et al.) nor very meaningful. With many hundreds or thousands of participants, we feel that the effect size (as measured by the increase in R^2) is a more appropriate indication of the influence of emodiversity than a *p* value.

This problem appears to have arisen, at least in part, as a result of a misunderstanding of the principles of a mathematical concept that has been imported from another field. As we have shown, Quoidbach et al.'s (2014) application of Shannon entropy is inappropriate, given the constraints imposed by the nature of the measures of emotional experience being used, such as the fact that a fixed number of emotions are measured, corresponding to a fixed list of "species," and a limited number of responses are allowed, corresponding to a limited range of possible population values. (This is not to suggest that, when applied to data obtained from more appropriate measures, Shannon entropy might not be useful in other areas of psychology or the social sciences more generally; see, for example, Vaquero & Cebrian, 2013). As one of us has pointed out previously (Brown, Sokal, & Friedman, 2013), researchers in psychology and other social sciences who wish to borrow concepts from the natural sciences or mathematics should ensure that they understand all of the conditions for the use of those concepts to be valid. The fact that a set of mathematical formulae (such as those for calculating Shannon entropy, but also those for performing ordinary least squares regression) can be applied to psychological data is no guarantee that the results that emerge from the application of these formulae will have any meaning in the real world.

In conclusion, we do not claim that the idea of emotional diversity is inherently devoid of any possible utility. However, in order for such utility to be demonstrated, it will be necessary to identify ways of operationalizing and measuring this construct that are not compromised by mathematical artifacts and statistical confounds. In its present form, Quoidbach et al.'s (2014) construct of emodiversity does not meet this standard.

References

Berbaum, K. S., Dorfman, D. D., Franken, E. A. Jr., & Caldwell, R. T. (2002). An empirical comparison of discrete ratings and subjective probability ratings. *Academic Radiology*, *9*, 756–763. http://dx.doi.org/10.1016/S1076-6332(03)80344-6

Brown, N. J. L., Sokal, A. D., & Friedman, H. L. (2013). The complex dynamics of wishful thinking: The critical positivity ratio. *American Psychologist*, *68*, 801–813. http://dx.doi.org/10.1037/a0032850

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, *34*, 35–46. http://dx.doi.org/10.1177/001316447403400105

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, *175*, 526–536. http://dx.doi.org/10.1097/00005053-198709000-00004

DeSoto, K. A. (2016). Computerized methods for collecting confidence ratings: Task influences in patterns of responding. Manuscript submitted for publication. Preprint available from https://peerj.com/preprints/1432v1/

Fredrickson, B. L. (2013). Positive emotions broaden and build. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47), pp. 1–53. Burlington, VT: Academic Press.

Kristjánsson, K. (2003). On the very idea of “negative emotions.” *Journal for the Theory of Social Behaviour*, *33*, 351–364. http://dx.doi.org/10.1046/j.1468-5914.2003.00222.x

Magurran, A. E. (2004). *Measuring biological diversity*. Malden, MA: Blackwell.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*, 858–865. http://dx.doi.org/10.3758/BF03194112

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. http://dx.doi.org/10.1037/h0043158

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research*, *39*, 303–328. http://dx.doi.org/10.1207/s15327906mbr3902_7

Quoidbach, J., Gruber, J., Mikolajczak, M., Kogan, A., Kotsou, I., & Norton, M. I. (2014). Emodiversity and the emotional ecosystem. *Journal of Experimental Psychology*: *General*, *143*, 2057–2066. http://dx.doi.org/10.1037/a0038025

Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality & Social Psychology*, *83*, 198–215. http://dx.doi.org/10.1037/0022-3514.83.1.198

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423. http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x

Solomon, R. C. (2002). On the very idea of "basic emotions." *Journal for the Theory of Social Behaviour*, *32*, 115–144. http://dx.doi.org/10.1111/1468-5914.00180

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the *p* value is not enough. *Journal of Graduate Medical Education*, *4*, 279–282. http://dx.doi.org/10.4300/JGME-D-12-00156.1

Vaquero, L. M., & Cebrian, M. (2013). The rich club phenomenon in the classroom. *Nature Scientific Reports*, *3*, 1174. http://dx.doi.org/10.1038/srep01174

Watson, D., & Clark, L. A. (1994). *The PANAS-X*: *Manual for the Positive and Negative Affect Schedule-expanded form*. Ames: The University of Iowa.

Table 1

Hierarchical linear models of positive, negative, and global emotion and emodiversity predicting depression from Quoidbach et al.'s (2014) Study 1.

Model number	Variables	Model R^2	β	*SE*	VIF
P.1		.381			
	Positive Emotion		−.617	0.017	1.000
P.2		.394			
	Positive Emotion		−.485	0.026	2.298
	Positive Emodiversity		−.175	0.026	2.298
P.3		.397			
	Positive Emotion		−.396	0.010	5.939
	Positive Emodiversity		−.364[c]	0.018	18.484
	Interaction		−.143[r]	0.013	9.288
N.1		.460			
	Negative Emotion		.678	0.004	1.000
N.2		.460			
	Negative Emotion		.682[s]	0.007	2.685
	Negative Emodiversity		−.005[s]	0.007	2.685
N.3		.466			
	Negative Emotion		.863[s]	0.011	8.149
	Negative Emodiversity		−.196[s]	0.012	8.748
	Interaction		−.143	0.007	3.409
G.1		.592			
	Positive Emotion		−.402	0.004	1.219
	Negative Emotion		.507	0.004	1.219
G.2		.595			
	Positive Emotion		−.356	0.005	2.061
	Negative Emotion		.553[c]	0.005	2.047
	Global Emodiversity		−.070[s]	0.005	1.961

Notes:

[s] : Denotes a coefficient that has either a greater magnitude or a different sign compared to the corresponding zero-order correlation, indicating that suppression has occurred.

[c] : Denotes a coefficient that has increased in magnitude from the previous step, indicating that some form of confounding has occurred.

[r] : The results in this table were calculated with SPSS. When we performed the same regression in R, we obtained a value of −.094 (*SE* = 0.008) for this coefficient. Only one of the other values in this table differed by more than .002 between SPSS and R, namely the coefficient for the interaction in section N.3, where R reported a value of −.137, a difference of 0.006.

Supplemental Information

This supplemental information document is divided into three parts. Part 1 provides detailed descriptions and results of the numerical analyses of the effects of bounded richness and evenness on Quoidbach et al.'s (2014) construct of emodiversity. Part 2 presents some extended footnotes to correspondingly-named sections in the main article. Part 3 contains some specific standalone points that we omitted from the main article due to lack of space, and because they were less central to our argument.

Part 1: Numerical analyses of the effects of bounded richness and evenness

Table S1 shows the range of emodiversity for each possible number of zero-scored responses (corresponding to *Never* experiencing a particular emotion) on three emotion-frequency measures. The first two measures have nine and 20 items, respectively, with possible responses in the range 0–4; the third measure has nine items with possible responses ranging from 0–100. The numerical effects of the severe constraints on the upper bounds on both richness and evenness imposed by Quoidbach et al.'s (2014) formulation of emodiversity can readily be seen in the rather small ranges between the minimum and maximum possible emodiversity values for any given number of non-zero (*Never*) responses. In each case, the lowest emodiversity is obtained when half of the non-zero responses are the lowest possible (i.e., 1) and the other half are the highest possible (i.e., 4 or 100, depending on the scale length); meanwhile, the highest emodiversity occurs when all of the non-zero responses are identical, regardless of which of the possible values in the range 1–4 (or 1–100) these non-zero responses all take.

It can be readily seen that, once the number of zero (*Never*) responses is established, the possible variation in emodiversity is severely constrained, as indicated by the "Ratio" entries in Table S1, which show the ratio between the maximum and minimum possible emodiversity values for each number of zero responses. The maximum/minimum ratio increases only slightly as the number of zero responses increases; for the 20-item scale, the ratio progresses from a minimum of 1.069 to a maximum of 1.385 (in the case of a participant who reports *Never* experiencing 18 out of 20 emotions, plus one *Rarely* and one *Most of the time*, which might well be considered a severely atypical pattern of responses), until the entire process breaks down with 19 or 20 zeroes. When the range of possible responses is expanded from 0–4 to 0–100, the ratio between maximum and minimum emodiversity for any given number of non-zero responses becomes somewhat larger; however, for the minimum evenness (and, hence, lowest emodiversity) to be obtained, participants would need to alternate between responses of 1 and 100 for every pair of emotions on the scale, which again would seem to be a highly unusual pattern, possibly indicative of acquiescence or boredom in participants. Thus, the range of possible "legitimate" emodiversity values needs to be further reduced to take into account that reports of either maximum or minimum emodiversity may be more likely to reflect unthinking response patterns than a sincere reflection of participants' experiences.

A further problem whose genesis can be glimpsed in Table S1, and which we were able to observe more concretely in Quoidbach et al.'s (2014) empirical studies, is the likely presence of a high degree of skewness in emodiversity, because of the strong effect of a single experience of an emotion (moving the response from *Never* to *Rarely*), compared to that of other responses. For example, Quoidbach et al.'s Table 3 shows that in Study 2 the maximum positive emodiversity score was 0.47 standard deviations (SDs) above the mean, whereas the minimum score was 13.1 SDs below the mean; indeed. Quoidbach et al. reported (p. 2062) that they applied a transformation to the emodiversity data in this study to address this skewness. In Study 1 the skewness was less dramatic, in that the maximum positive emodiversity was 0.72 SDs above the mean and the minimum was "only" 4.32 SDs below the mean; nevertheless, the overall skewness for positive, negative, and global emodiversity in Study 1 were −2.52, −0.92, and −1.54, respectively. We investigated the effects of attempting to correct this skewness by squaring (or raising to an even higher power) these emodiversity variables. However, when we did this, the correlation of these new transformed variables for positive and negative emodiversity with their respective emotion measures became even higher than before (e.g., for negative emodiversity, the correlation with negative emotion increased from .79 to .86 when emodiversity was squared). Hence, it seems that this high skewness is likely to be another suboptimal aspect of emodiversity data that researchers will have to cope with in their analyses.

Table S1

Minimum and maximum possible emodiversity values, and the ratio of maximum to minimum, for each possible number of zero-scored responses on three different possible scales with different lengths and scoring ranges.

	9-item scale (scored 0–4)			20-item scale (scored 0–4)			9-item scale (scored 0–100)		
Zeroes	Minimum	Maximum	Ratio	Minimum	Maximum	Ratio	Minimum	Maximum	Ratio
0	1.988	2.197	1.105	2.803	2.996	1.069	1.456	2.197	1.510
1	1.887	2.079	1.102	2.744	2.944	1.073	1.442	2.079	1.442
2	1.733	1.946	1.123	2.698	2.890	1.071	1.172	1.946	1.660
3	1.599	1.792	1.121	2.632	2.833	1.077	1.154	1.792	1.552
4	1.390	1.609	1.158	2.580	2.773	1.075	0.776	1.609	2.074
5	1.194	1.386	1.161	2.505	2.708	1.081	0.749	1.386	1.852
6	0.868	1.099	1.266	2.446	2.639	1.079	0.110	1.099	9.978
7	0.500	0.693	1.385	2.361	2.565	1.087	0.056	0.693	12.479
8	0.000	0.000	n/a	2.292	2.485	1.084	0.000	0.000	n/a
9	0.000	0.000	n/a	2.192	2.398	1.094	0.000	0.000	n/a
10				2.110	2.303	1.091			
11				1.988	2.197	1.105			
12				1.887	2.079	1.102			
13				1.733	1.946	1.123			
14				1.599	1.792	1.121			
15				1.390	1.609	1.158			
16				1.194	1.386	1.161			
17				0.868	1.099	1.266			
18				0.500	0.693	1.385			
19				0.000	0.000	n/a			
20				0.000	0.000	n/a			

Notes:

1. The possible emodiversity values for the 9-item (scored 0–4) scale with 0, 1, 2, etc., zero-scored items are identical to the corresponding items for the 20-item (scored 0–4) scale with 11, 12, 13, etc. zero-scored items. More generally, the range for an N-item scale (N < 20) can be obtained by taking the last N items of the 20-item scale.
2. The alternating increase and decrease in the ratio as the number of zeroes increases (particularly noticeable for the 9-item scale scored 0–100) represents the fact that minimum emodiversity increases by a smaller amount when the number of non-zero items on the scale becomes odd (when a value of 1 is added to the list of non-zero item scores) than when it becomes even (when the maximum value or 4 or 100 is added to the list).

Part 2: Extended Footnotes

Notes for "**Limitations of the analogy with biodiversity**"

Richness vs. Evenness: There is some debate among biologists as to whether richness and evenness should be considered separately, because they are typically closely related empirically (Stirling & Wilsey, 2001). While specific individual measures of both richness and evenness exist, the Shannon entropy formula—sometimes referred to as the Shannon-Wiener index—functions as a composite measure: Once the Shannon-Wiener value (denoted by H') for the biodiversity of a community has been established, an increase in either richness or evenness will produce a larger H' value. This ambiguity could be considered a weakness of attempts to repurpose Shannon's original concept beyond its originally intended field of communication. In Shannon's (1948) model, the practical outcome of either a greater number of different characters in the message ("richness") or a greater variety in the distribution of those characters ("evenness") is the same, namely an increase in the number of bits required to uniquely encode the message. In other words, when applied to electronic communications, richness and evenness are two sides of the same coin, whereas in biology (and emodiversity), this relation is considerably more complex.

Taxonomy of emotion scales: The nine positive emotions measured by the mDES, used by Quoidbach et al. (2014) in their Study 1, are *alertness*, *amusement*, *awe*, *contentment*, *gratitude*, *hope*, *joy*, *love*, and *pride*. In their Study 2, Quoidbach et al. used a different, unnamed measure, with ten positive emotions: *amusement*, *awe*, *contentment*, *enthusiasm*, *gratitude*, *happiness, interest*, *joy*, *pride*, and *serenity* (of which only six are common to this measure and the mDES). Had Quoidbach et al. instead taken their set of emotions from the PANAS (Watson, Clark, & Tellegen, 1988), which certainly seems plausible—for example, the filename of the "emodiversity calculator" spreadsheet that we downloaded from the `emodiversity.org` website on October 28, 2015 was "`Emodiversity Calculator 20 emotion items (like PANAS).xlsx`"—their list would have been even more different from the mDES. The PANAS uses adjectives rather than nouns to ask participants how they are (or were) feeling; its list of 10 positive emotions consists of the words *active*, *alert*, *attentive*, *determined*, *enthusiastic*, *excited*, *inspired*, *interested*, *proud*, and *strong*. Thus, the mDES and PANAS have only two positive emotions (*alertness* and *pride*) in common. It is difficult to imagine how one could reliably compare emodiversity values across populations using such different instruments. Yet, the emodiversity calculator spreadsheet invites researchers to use any emotion measure of their choice, as long as the scale is zero-based. (Note that a base of zero for the responses is a mathematical requirement of the way the emodiversity formula is constructed; any existing scale that uses a response format that starts above zero, such as 1–7, will need to be rebased to zero in order to be used to calculate emodiversity.)

A further complication is that emodiversity values that have been derived from two measures with different numbers of items are not commensurate. The range of emodiversity values for a 10-item scale (maximum emodiversity 2.30) is different from that of an 18-item scale (maximum emodiversity 2.89). Thus, unless a decision is made to standardize on a definitive number of items, there is no prospect of a universal scale of values of emodiversity; any given individual will have a 10-item emodiversity score, an 18-item score, and so on. This represents a further departure from the application of Shannon entropy to communication or biodiversity; in both of those cases, any message or community has exactly one entropy value, and these values can be readily compared across messages or communities.

Notes for "**Empirical Issues**"

Need for stepwise reporting. Quoidbach et al. (2014) reported the results of their multiple regressions only at the final step, with all variables included. But in the presence of correlated predictors, without an indication of the evolution of the model as variables are added, it is almost impossible to evaluate whether the focal variables of the study are really those that are driving the observed effects. Had Quoidbach et al. followed this recommendation, reporting the extra variance explained at each step, the limited power of emodiversity to explain unique variance (beyond the corresponding measure of emotions) in both of their studies would have been more readily apparent. (We note that Quoidbach et al. did report, in the summary paragraph for their Study 1, that emodiversity explained only about 1% of the variance in depression.)

No apparent checks for automatic responding. We noted in our main article that the highest and lowest values of emodiversity are to be found in cases of extreme response patterns. The participants in Quoidbach et al.'s (2014) Study 1 were recruited through a TV advertisement that appeared in a popular show dedicated to happiness; this advertisement invited viewers to visit the show's website in order to participate. In addition to the risk of sampling bias, the relatively casual nature of this recruitment process might be expected to include a number of people who might not necessarily spend a lot of time carefully considering their responses. Indeed, of Quoidbach et al.'s 41,723 participants, 637 reported identical frequencies of experience of all nine positive emotions (288 *Never*, 349 some non-zero frequency), while 1,298 (1,220 *Never*, 78 non-zero) reported identical frequencies of experience of all nine negative emotions. However, it does not appear that any participants were excluded from Quoidbach et al.'s analyses purely on the basis of identical responses.

<u>Part 3: Some further topics</u>

Distortions introduced by the time period under study

The capacity of any of the instruments used by Quoidbach et al. (2014) to measure emodiversity depends on the time period for which participants are invited to report their emotional experiences. This is because, of the possible responses to the scale items, one (*Never*) is an absolute number (i.e., the emotion was experienced on a total of zero occasions during the entire period in question, whatever the length of this period), whereas the others are expressed in terms of an informal proportion of the time available. For example, consider what would happen if the word "week" in the (presumed) question "How often did you experience each of these emotions in the past week?" were to be replaced with the word "month." Assuming that a person's experiences of emotions are fairly stable over time, their relative proportion of responses from *Rarely* through *Most of the time* ought to be similar for any given time period; that is, someone who, over the course of a typical week, is alert *Most of the time* and only *Rarely* embarrassed will probably report experiencing those emotions with similar relative frequencies over the course of a month. Indeed, in order for Quoidbach et al.'s claims about the ability of emodiversity scores measured over any given week (which we presume was the time period used in these authors' empirical studies) to predict long-term mental and physical health outcomes to be valid, such an assumption about the temporal stability of the frequency of emotional experiences would seem to be not just plausible, but necessary. Otherwise, Quoidbach et al.'s inferences about health outcomes over the course of months or years from a single week's emodiversity scores would appear to be potentially subject to a great deal of random variation depending on the week in question.

In contrast, when the period under consideration is extended from a week to a month (or longer), the number of *Never* responses is likely to be reduced, simply because more time is available for each emotion to have been experienced at least once. Perhaps Dave has not been embarrassed in the past week, but he might have had an awkward moment three weeks ago, thus causing him to respond that he experienced embarrassment *Rarely* (versus *Never*) when the time period is expanded to a month. The importance of this point is illustrated when we examine the influence of zero-coded (*Never*) responses on the calculation of emodiversity. In the data set for Quoidbach et al.'s (2014) Study 1, we found a correlation of −.96 between the count of *Never* responses for each participant to either the positive or negative mDES subscale and the respective measure of emodiversity. (This correlation would be somewhat lower for a 0–100 response format, but only to the extent that participants use a wider range of response values, which—as we mentioned in our main article—appears to be unlikely in practice.) This remarkably high correlation provides a further illustration of the point that we made in our main article regarding the limited range of possible emodiversity values (indeed, it is a mathematical corollary of that limited range). Once the number of *Never* responses is established, emodiversity is almost completely static, so that increasing one's experience of an emotion beyond *Rarely* has almost no effect (and what effect it does have is as likely to be negative as positive); yet, the number of *Never* responses depends on the essentially arbitrary choice of time period used for the measure of emotions. A measure that asked people how many times they had experienced each emotion in the past year—which might be highly valid in an investigation of people's long-term emotional experiences—would likely reveal that many of them had almost the same emodiversity, as most people would not be able to give a response of *Never* to the majority of the items.

Forced responses to all scale items

Shannon's (1948) index of entropy quantifies the diversity among (only) the observed characters in a message. For example, the one-word message *cataract* contains three observations of the letter *a*, two each of *c* and *t*, and one of *r*, and has a Shannon entropy of

1.91[1]. The fact that the other letters of the alphabet do not appear in this message is irrelevant, as is the length of the alphabet. Similarly, when the Shannon-Wiener index is used in field biology, only the observed species are of interest; the fact that no aardvark or zebra was sighted during a field study of a Welsh bog does not affect the biodiversity index of that community. By definition, there is no such thing as an observation of zero occurrences of a character, or zero members of a species, in these situations. In contrast, the structure of the psychologist's questionnaire—which is equivalent to taking along a pre-printed checklist of possible species, rather than a blank sheet of paper, to observe a biological community—means that there is a difference between the absence of any report concerning a particular phenomenon, and a concrete statement that the phenomenon did not occur. The psychological equivalent of the absence of the letter *q* in Shannon's message, or of the failure to observe any tigers in a field study in biology, is not a report by the participant that an emotion was never experienced; rather, it is the *absence of a response* to the item concerning that emotion. These two situations are clearly distinct. For example, when the mDES is used to calculate a participant's mean level of positive emotions, a response of *Never*—coded as zero—to a particular emotion item has an effect on the participant's mean positive emotions score, by increasing the denominator (i.e., the total number of responses), in a way that a failure to respond—which would simply be ignored in the calculation of the mean—would not.

[1] This value is calculated using binary (base 2) logarithms, as used by Shannon (1948). Binary logarithms are appropriate in the case of communication systems because one is typically interested in the number of bits that a coded message will occupy. Natural logarithms, as used by Quoidbach et al. (2014), are arguably more appropriate for other applications of Shannon's formula. When natural logarithms are used, the Shannon entropy of the message *cataract* is multiplied by ln(2), or 0.6931, to give 1.32.

Quoidbach et al. (2014) therefore needed to handle the separate possibilities of a missing response and an explicit zero-coded response[2]. However, their choice to treat a response of *Never* as making no contribution to emodiversity—by assigning a value of zero for the corresponding ($p_i \times \ln p_i$) term in Shannon's formula[3]—effectively transformed *Never* into a missing response. In turn, this meant that a decision had to be made about how to handle cases where the participant did in fact fail to respond to one or more items, because to treat a genuine missing response as being identical to a response of *Never* would represent a serious distortion of the meaning of the items composing the scale. Quoidbach et al. did not discuss this problem in their article, but from an examination of their examples, the data set for their Study 1, and their online emodiversity test at `http://www.emodiversity.org/`, it appears that their solution was to *force* participants to answer every question on the mDES or equivalent measure. For example, 6,611 participant records in the data set for Quoidbach et al.'s Study 1 have no data for any of the items on the nine-item MADR-S depression scale, but a further 848 records have data for between one and eight items, suggesting that no forced choice was imposed for this scale. In contrast, 5,879 participants have no data for the mDES, but none of the records in the data set have partially complete responses for this scale. This suggests to us that responding to all items was mandatory, probably being enforced by the computer system used for the survey. (We were unable to check the equivalent numbers for Study 2, as the data set has not been made public.)

---

[2] The nature of the $p_i \times \ln p_i$ formula for calculating the contribution of each item to emodiversity means that the scale being used must be (re-)anchored at zero, as noted by Quoidbach and colleagues in their emodiversity calculator spreadsheet.

[3] Although the logarithm of zero is undefined, the ($p_i \times \ln p_i$) term for an item whose value (and, hence, p_i term) is zero, is defined (with a value of zero).

In summary, a decision to measure emodiversity would seem to require researchers to take a rather Procrustean (and, potentially, bias-inducing) approach to participants: Only those who are willing and able to provide responses to every item on a scale measuring the frequency of experienced emotions are candidates for inclusion in calculations of emodiversity. However, forcing participants to respond in this way potentially introduces a number of psychometric problems (Ray, 1990). It also implies that emodiversity will be difficult to measure in pencil-and-paper settings, where participants always have the opportunity to not answer a particular item (whether deliberately or not); to exclude those who consciously (or accidentally) omit one or more responses might be a source of bias. Furthermore, the more items in the emotion measurement scale being used, the greater the chance that participants will omit one or more responses and, hence, that their results will not be used at all. This potentially creates a perverse incentive for researchers to use shorter measures, which will not only be less reliable for measuring the emotions that they are supposed to measure, but will also allow for even less variance in emodiversity.

The values of emodiversity are produced by the use of invalid arithmetic

Quoidbach et al. (2014) stated that they obtained the p_i components of the Shannon entropy formula by "divid[ing] the number of times an individual experienced [an emotion] by the total number of times she experienced all types of emotions" (p. 2058). However, it is not correct to say that a measure such as the mDES gives an indication of the "number of times" each emotion was experienced. Instead, an arbitrary integer value is applied to a subjective verbal indicator of the frequency with which each emotion was experienced, ranging from *Never* (scored as 0) to *Most of the time* (scored as 4). But, even if it can be shown empirically that multi-item ordinal (Likert) scales typically behave empirically like interval data when summed and averaged (Carifio & Perla, 2007), the act of dividing an individual number from such a scale by any other number (in this case, the equally-arbitrary total of all the emotion-frequency scores) is a meaningless operation, and subsequently multiplying the result of that operation by its own logarithm merely compounds this error. All that can be said about the results of such a numerical manipulation is that the p_i terms corresponding to more frequent experiences of each emotion will be larger than those corresponding to less frequent experiences, but the magnitude of this relation is essentially arbitrary, and will vary depending on the frequency of experience of other emotions endorsed *by the same participant*. Even if the range of possible emotion-frequency responses were to be extended to 0–100, as discussed earlier, it would seem to be difficult to defend the idea that this creates an interval scale, with a value of 60 somehow representing experiencing an emotion "twice as often" as a value of 30.

The equal desirability of all emotions on the scale being used

One of the limitations of Quoidbach et al.'s (2014) operationalization of emodiversity is that it places equal weight on all of the emotions on the scale that is used (which, as we noted in our main article, is already a somewhat arbitrary selection from the possible taxonomy of emotions). Yet, just as not all forms of biological evenness are necessarily desirable (a hectare of African savanna that currently contains an equal biomass of wildebeest and lions is likely to undergo some kind of upheaval in the near future), so it is not clear that it is beneficial for people to experience all emotions, even those typically considered "positive," with equal frequency. For example, while it might be good for people to experience alertness or contentment *Most of the time*, it might be better for the stress levels of their family and friends if these same individuals experienced pride or awe only *Some of the time*. Yet, Quoidbach et al.'s theoretical approach—which, for any given level of richness, rewards evenness with a higher emodiversity score—implies that all emotions (at least, those of a given valence) are created equal, and that experiencing them all with the same frequency—whatever that frequency might be—is a good thing.

The effect of adding quadratic terms to the regression models

Cortina (1993) noted that when predictors are highly correlated, the effects of their interaction term can become confounded with the effects of higher order (typically quadratic) terms of each predictor. We therefore tested the effects of adding the square of emotion and emodiversity terms to the principal regression models for positive and negative emodiversity in Quoidbach et al.'s (2014) Study 1, using Cortina's (1993, p. 918) three-step process.

For positive emotion and emodiversity, R^2 was .391 at step 1 (two predictors, cf. Table 1, section P.2 in our main article), .397 at step 2 (after adding the two quadratic terms), and .400 at step 3 (after adding the interaction). The added variance explained by the interaction (R^2 increase of .003) for positive emotion and emodiversity was therefore the same as in Quoidbach et al.'s (2014) basic model.

For negative emotion and emodiversity, R^2 was .460 at Cortina's (1993) step 1, .465 at step 2, and .466 at step 3 (after adding the interaction), so that the added variance explained by the interaction after accounting for the confounding effects of the quadratic term (an increase in R^2 of .001) was only approximately one-sixth of what was observed by Quoidbach et al. (2014) in their basic model for negative emotion and emodiversity when the interaction term was added. (As sections N.1 and N.2 of Table 1 in our main article show, the interaction term was the *only* source of extra explained variance in the regressions for negative emotion and emodiversity; adding negative emodiversity on its own to a regression containing only negative emotion as a predictor produced *zero* increase in explained variance, to three decimal places.)

References

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*, 106–116. http://dx.doi.org/10.3844/jssp.2007.106.116

Quoidbach, J., Gruber, J., Mikolajczak, M., Kogan, A., Kotsou, I., & Norton, M. I. (2014). Emodiversity and the emotional ecosystem. *Journal of Experimental Psychology*: *General*, *143*, 2057–2066. http://dx.doi.org/10.1037/a0038025

Ray, J. J. (1990). Acquiescence and problems with forced-choice scales. *Journal of Social Psychology*, *130*, 397–399. http://dx.doi.org/10.1080/00224545.1990.9924595

Stirling, G., & Wilsey, B. (2001). Empirical relationships between species richness, evenness, and proportional diversity. *The American Naturalist*, *158*, 286–299. http://dx.doi.org/10.1086/321317

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063